APPENDIX E

LAND USE AND DEMOGRAPHICS DEVELOPMENT

Contents

E.1 INTRODUCTION

The methodology, analysis, and forecast population for the MASE EA was developed from January 2005 to March 2005. To support noise impact analysis in the EA, future year population projections to the Census Block (CB) level of detail was required. The forecast effort was required since the 2000 US Census information is not considered to be current enough to support a reasonable analysis of the potential noise impacts for the EA alternatives on the population within the Environmental Study Area. This analysis provides CB forecasted population information for all CBs in the Environmental Study Area for years 2004, 2006 and 2011.

Since the Environmental Study Area also consists of a small part of Southern Ontario bordering Michigan, forecasting was carried out for this area as well. In this area, Canadian Census geography was used. The smallest area type available for analysis was the Canadian Census Dissemination Area (DA). Thus in the US part of the Environmental Study Area, CBs were analyzed while in the Canadian part of the Environmental Study Area, DAs were analyzed. Overall, the Environmental Study Area consisted of 172,526 CBs and 716 DAs.

The initial forecasting effort focused on only Total Population by CB or DA. Due to the limited demographic information available at the CB level of geography, population projections were initially collected only at the Census Block Group Level (BG). Subsequently, HNTB dispersed demographic information collected at the BG level down to the CBs located within each BG.

E.2 DATA

The major items of US data utilized in this report consist of:

- Census Geography
- Census Data
- Population Projection Data
- Land Use Geography

Due to the limited availability of Canadian data it is described separately.

E.2.1 Census Geography

Census Geography was obtained from Environmental Systems Research Institute (ESRI). This consisted of 2000 CB and BG boundary shapefiles. Geography, in the context of this appendix, refers to the physical boundaries of the CB or BG. Since the population projection data assumes no change in Census Geography from the year 2000 geography, these shapefiles are consistent with all the population projection data.

A key element of these shapefiles is that they contained the CB number as an identifier for each CB shape. In addition it is possible to derive the BG from the CB identifier. Thus, the population projection data can be linked to the Census geography through these identifiers. Consequently, the population projection data can then be linked to the original Census Data.

E.2.2 Census Data

Census Data consisting of population by CB for year 2000 was also obtained from ESRI. This data included total population for each CB, which was a key ingredient in the analysis.

E.2.3 Population Projection Data

The population analysis for the EA required population projection data for the years 2004, 2006 and 2011.

Applied Geographic Solutions (AGS) was selected as the vendor for the population project data. AGS was able to provide total population data at the BG level for the years 2004, 2009, and 2014. Thus, the data needed to be to the CB level. Also, it needed to be interpolated for the study years 2006 and 2011.

E.2.4 Land Use

Land use is a critical element for successful dispersion of the BG data to the CB level. The key issue is distinguishing residential from non-residential land use so that correct population predictions can be made as to where the population is likely to be located in the future. The Environmental Study Area was divided into three sub-areas for the purposes of land use: two 10-nautical-mile areas around DTW and CLE, and the rest of the Environmental Study Area. This is shown in Table E-1.

Table E-1
Land Use Data Sources by Study Sub-Area

Study Sub-Area	Land Use Data Source	Number of Census Blocks
10-nautical-mile around CLE	Cuyahoga County	8,987
10-nautical-mile around DTW	Southeastern Michigan Council of Governments	13,468
Michigan	National Land Use Cover Dataset (NLCD)	71,815
Ohio	NLCD	78,256
Total		172,526

The local land use data sources consisted of land use data obtained from Cuyahoga County and Southeastern Michigan Council of Governments (SEMCOG) for Cleveland and Detroit respectively. The local data is a significant enhancement over the NLCD data and provides a much more accurate picture of current land use.

E.2.5 Canadian Data

Canadian data suffered from limited availability. The only data available was at the Canadian Census Dissemination Area (DA) level. In addition the only geographic information useful for this study was the centroid of each DA. The population data was obtained from Tetrad Inc. which has based the data on MapInfo Canada and Statistics Canada. The Canadian Census was held in 2001. Population projection data was available for 2004, 2007, 2009, and 2014. Thus, interpolation was used for 2006 and 2011 similar to the US data. No land use data was available and none was needed since a dispersion of population to smaller sized geography was not possible for Canadian data due to the lack of smaller scale geography. A total of 716 DAs from southern Ontario were analyzed for the study.

E.3 ANALYSIS

This section consists of a description of analysis steps, key algorithms and data transformations utilized in the analysis.

E.3.1 Steps

US Region

1 **Source Data Metadata Completion:** Obtain all source data from the variety of sources identified in the data section. Complete metadata on all source data (where available). Conform to Federal Geographic Data Committee (FGDC) format if it is geographic data. This ensures that we can clearly document our data sources.

2 **Source Data Re-projection:** Re-project all source data if necessary so that we are dealing with one spatial projection for all subsequent analysis.

3 **Source Data Editing:** Perform limited editing on source data. Because of the size of the Environmental Study Area and the nature of the data, it is not possible to do large scale data editing and quality checking. However, certain obvious flaws can be removed. For instance, a few obvious land use data errors around the airports were corrected. These edits were documented.

4 **Geographical Dataset Preparation:** Three base datasets were prepared. These consisted of the 10-nautical-mile sub-area around DTW, the 10-nautical-mile sub-area around CLE, and the whole Environmental Study Area encompassing these areas as well. The base dataset preparation was similar for each dataset. The only difference was that local land use data was used for the two close-in areas and NLCD land use data was used for the overall Environmental Study Area. The preparation step consisted of clipping the geographical areas to get just the needed datasets of census and land use geography.

5 **Land Use Data Transformation:** Land use classifications came in different varieties based on data source. These were all conformed to a two class scheme – residential and non-residential based on the descriptions of each original land use class. For instance, if the original classes of land use were agricultural or industrial, they were all re-classed as non-residential.

6 **Census Block Residential Area Calculation:** The land use data is spatially merged into the CB geographies to get residential land use area per CB.

7 **Census Block Group Geography and Population Forecast Data Processing:** The BG is extracted from the CB identifier and all CBs belonging to a BG are merged spatially to form BG geographies. In addition, relevant data such as residential land area is aggregated from the individual CBs to get data for the BGs. The BG geography is then attached to the initial population forecast data from AGS using the BG identifier as the relational key.

8 **Forecast Population Dispersal Block Group to Census Block:** The population BG forecast data is dispersed down to the CB based on an enhanced 'Equal Area Distribution with Control Zones' algorithm. The algorithm is described later on in this section.

9 **Interpolation Forecast Years to Study Years:** Since the available forecast data (2004, 2009, and 2014) did not match the study year needs (2004, 2006, and 2011) it was interpolated using an interpolation algorithm described elsewhere in the document.

10 **Final Data File Preparation:** The final processing of the files consisted of removing the 10-nautical mile areas around DTW and CLE from the remainder of the Environmental Study Area, since the 10-nautical areas had their own forecasted data using better land use. The individual CB centroids (i.e., a geographical point in the middle of the CB) were calculated. Finally, the data was exported to a DBF file format, file descriptions were attached, and data and descriptions were transmitted to Metron Aviation for the noise analysis.

Canadian Region

The Canadian data analysis steps consisted of forecast year interpolation. The data was available for 2004, 2007, 2009, and 2014. It was interpolated for 2006 and 2011. In addition, the base census year for Canada is 2001 while it is 2000 for the US Since data needed to be delivered for 2000, 2004, 2006, 2011 for both US and Canadian areas within the Environmental Study Area, the 2000 year data was additionally extrapolated for the Canadian data.

E.3.2 Algorithms

Population Dispersal Block Group to Census Block

Given the needs of the EA, the challenge is to disperse population data available at a larger geographical area to a number of smaller geographical units within it. This is necessary since the population forecast data is not available for the smaller unit (i.e., CB) but is only available for the larger unit (i.e., BG). Two methods were reviewed for population dispersal: the (1) Equal Area and the (2) Equal Area with Control Zones. The Equal Area with Control Zones was recommended for population dispersal. As a part of this project, HNTB's analysis has indicated that the dispersal technique could be further improved by making some changes to the methodology that had been used in previous projects. These are discussed below.

The Equal Area method simply takes a population at the BG level and prorates it to the individual blocks based on their individual areas. The problem with this approach is that it does not take into account existing land use. Thus, if the largest area block within a BG happens to be a park, the Equal Area method would still assign the largest percentage of population to it. This was the reason that this methodology was not used. Instead, the Equal Areas with Control Zones method was selected. This method takes into account current land use and distributes population based on residential land use area not the total area of each block. Thus, in the above example although the BG has a large total area, it would have a small or none residential land use area. Therefore, it would get very little or no population distributed to it.

HNTB's analysis indicates that the Equal Areas with Control Zones method needs to take into account current population densities in addition to current land use, especially for low population growth areas such as inner cities. A conservative option is to only distribute the change in population based on residential land use. This method is called Population Change Dispersal using Equal Areas with Control Zones. The Population Change Dispersal method now distributes just the **change** (not the total) in the BG's population from the previous period according to existing land use. Thus, if a BG is forecasted to undergo no population change, the individual CBs within the BG will keep their current population. This technique results in smooth transitions in population within BGs. It may result in an over-estimation of population for very high growth areas. This is not an issue for the current Environmental Study Area which is not forecasted to see major population growth. In any case, very high growth areas would tend to see a change in current land use as well and would require a technique to project land use change within individual BG.

Forecast Year Interpolation

Forecast year interpolation was required since the data obtained from the vendors did not conform to the study year needs. The technique used for interpolation is described below.

Since 4 data points were available in all cases (for each individual CB), a cubic polynomial fit of the form:

$$ax^3 + bx^2 + cx + d$$

would pass through all the known points exactly. A Least-Square fit using this form of the equation was applied to each CB's data, resulting in a unique prediction equation (with known a, b, c and d values) for each CB. The unknown years were then calculated using this equation.

This form of interpolation works quite well when the data requiring interpolation is close to a known point. In this case all unknown years were within two years of a known year. In addition, the rate of change in population for most CBs is not very large. Thus, this form of interpolation is a good approximation.

In the case of Canadian data, this algorithm was actually used to extrapolate one particular year for the year 2000 analysis. This was needed since the Canadian Census year was 2001 and we needed all data to conform to 2000 – the base year from which projections were made for US data. This extrapolation is not inaccurate since actual data is available for 2001 and there is not much difference from year to year. In addition, this is a base year for analysis as opposed to the forecast years which are all interpolations (instead of extrapolations).

E.4 CONCLUSION

The methodology utilized in this report is an enhanced Equal Area with Control Zones methodology. In brief, the enhancements consist of:

- **Block Group to Census Block** forecasted population dispersion algorithms were modified to take into account the low growth demographic environment of the Environmental Study Area.

- **Individual Study Year's** forecasts were interpolated. Due to limited data availability, the initial projections at the BG level were for years that did not exactly correspond with study year requirements. The data was interpolated for the study years.
- **Land Use Data** was obtained at a higher quality level close-in to the airports. Thus two types of land use data were utilized for the Environmental Study Area: a close in higher quality land use database for 10-nautical-mile radius region around DTW and CLE, and more generalized land use for the rest of the Environmental Study Area.

These methodology enhancements significantly improve the quality of the demographic forecasts developed for this EA.